

Name: Deniz Koyuncu  
PI Name: Bülent Yener  
Presentation preference: In person

Email: koyund@rpi.edu  
PI email: byener@gmail.com

## **Missing Value Knockoffs for Genome-Wide Association Studies**

Deniz Koyuncu<sup>1</sup>, Bülent Yener<sup>2</sup>

<sup>1</sup> Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute; <sup>2</sup> Department of Computer Science, Rensselaer Polytechnic Institute

Genome-Wide Association Studies (GWAS) aims to identify the genetic causes of complex traits and commonly measure the individual effect of the genetic markers. Although multivariate genome-wide models such as penalized regression and random forests are more suitable for capturing epistasis, they lack any control of the type 1 error. Model-x knockoffs, a recent solution, provides a framework for False Discovery Rate (FDR) control with a wide range of multivariate models. One limitation of model-x knockoffs is its lack of support for missing values which can occur in Single Nucleotide Polymorphism datasets. In this work, we first introduce an imputation algorithm designed to preserve the FDR control guarantee of the model-x knockoffs framework. Next, we introduce a specialized imputation procedure for latent variable models which maintain the theoretical guarantees in a computationally efficient manner. To test the developed methods, we study how the missing data amount affects the power and FDR in different experimental settings. Our simulations show that under different amounts of missing values, observations, and two different families of distributions, FDR control is possible under the knockoffs methodology even when missing values are imputed. Developed methods can be applied in different GWAS such as the ones studying the use disorders and addictive behaviors.